


```

Db 121 TTTSTTAAFOQRTTFHEASRKYNDVAKIYSINTVNMNGVASYCRCALEASDVGSCTS 180
OY 605 CPAGYIIDRDSSTCHSCPPTTLKAHQYGVACVPCGCGTNNKTIHSLCTYDCITFSRNT 664
Db 181 CPAGYIIDRDSSTCHSCPPTTLKAHQYGVACVPCGCGTNNKTIHSLCTYDCITFSRNT 240
OY 665 PRFTNPNFSALANTVTLAGGSPFTSKGLKYFHHFTLSLCGNOGRKMSCTONVTDLRIP 724
Db 241 PRFTNPNFSALANTVTLAGGSPFTSKGLKYFHHFTLSLCGNOGRKMSCTONVTDLRIP 300
OY 725 EGESGFSKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPL 784
Db 301 EGESGFSKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPL 360
OY 785 FHLESIGIPDVIFEFYRSDNVYQSSSGRSTTIRVRCSPQKTYPGSILLPCTGSDGICDC 844
Db 361 FHLESIGIPDVIFEFYRSDNVYQSSSGRSTTIRVRCSPQKTYPGSILLPCTGSDGICDC 420
OY 845 NFHFMESAAACPLCSVDYHAIVSSCAVGIQKTYVWREPKLSCGISLPPORVITCKT 904
Db 421 NFHFMESAAACPLCSVDYHAIVSSCAVGIQKTYVWREPKLSCGISLPPORVITCKT 480
OY 905 IDFWLKVGISACTCAILLTVLTCYFWMKNQKLEYKSKLVNNAATLKDCDLPADSCALIM 964
Db 481 IDFWLKVGISACTCAILLTVLTCYFWMKNQKLEYKSKLVNNAATLKDCDLPADSCALIM 540
OY 965 EGEDVEDDLIFTSKNHSLGR 984
Db 541 EGEDVEDDLIFTSKNHSLGR 560

```

RESULT 2

```

OY 096DP2 PRELIMINARY: PRT: 493 AA.
AC 096DP2:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2002 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CDNA FLJ131340 f1s, clone MESANI000035, weakly similar to major
DE surface-labeled trophozoite antigen precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
RA Murakami K., Kanehori K., Takahashi-Fuji A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,
RA Iisaga T.;
RT "NDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK055902; BAB1041.1; -.
DR InterPro; IPR001881; EGF_CA.
DR SMART; SM00179; EGF_CA; 1.
SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

```

Query Match 28.2%; Score 1550; DB 4; Length 493;
 Best Local Similarity 55.7%; Pred. No. 7.3e-118;
 Matches 272; Conservative 80; Mismatches 132; Indels 4; Gaps 4;

```

OY 231 HSEVLNRCNNVLYRTTFSSVTAKVPKPLVLRNIAITGVAAYSECPKPGTYADKOGSS 290
Db 6 YQVMAKSGNIIYWRITGLMGSKAVKPVLVKNITIEGVAIYSECPKPGTFGFSKPGSF 65
OY 291 FCKLCPANYSYKSGTGHOC-DPDKYSEKSSSCNVPRPACTDKDYFTYTHACDANGETO 349
Db 66 NCOVCPRNTYSEKGAKECIRCKDSDQSFEGSSSECTERPPCTTKDYFOIHFPCEDEGKTO 125

```

```

OY 350 LMYKNAPKICSEDLLEGAVKLPASGVKTHCPQCPNPGFEFTKNNSTOCPCYGSXNGS-DC 408
Db 126 IMKWIPEPICREDLTDALRPPSGEKKDCPCPCNPBFYNNSSSCHPCPGTFSDTKKC 185
OY 409 TRCPACTEPAVGEFKWMNTLPTNMETVLSGINFEEKMTGWEVAGDHIYTAAGASND 468
Db 186 RCPACTEPAVGEFKWMNTLPTNMETVLSGINFEEKMTGWEVAGDHIYTAAGASND 245
OY 469 EMILTYVGEFRPPOVMAVDENKEVARITTFEFTLCSVNCLEFVWVGNSTRNPVENW 528
Db 246 YLLNHLHIGFKEPPYS-MGATGSELGRITFEFTLCSADCVLEFVWDINRSTNVEWM 304
OY 529 KSKSGOSYTYIIEENTTSFTMAFQRTTFHEASRKYNDVAKIYSINTVNMNGVASYC 588
Db 305 GGTEKEQATTHIIFKNATFTFMAFQRTNOGDNRRFINDMKIYITATNADVASSC 364
OY 589 RCPALASDVGSSTSCPAGYIIDRDSSTCHSCPPTTLKAHQYGVACVPCGCGTKNN 648
Db 365 RACALGSFQSGSSVCPCPGHYIEKETNOCKECPPTVLSIHQVYKEACIPCGPSKNN 424
OY 649 KIHSLCYNCTSRNTPTFTNPNFSALANTVTLAGGSPFTSKGLKYFHHFTLSLCGNG 708
Db 425 ODHAVCYSDCFEYPERKENSLHDFSNLSSVGLMNGSPFTSKGYTFHFHFNISTYVGMKG 484
OY 709 RK-MSVCT 715
Db 485 RKMLSVPT 492

```

RESULT 3

```

OY 08R215 PRELIMINARY: PRT: 300 AA.
AC 08R215:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 32.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC022655; AAH22655.1; -.
FT NON_TER
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

```

Query Match 23.3%; Score 1282; DB 11; Length 300;
 Best Local Similarity 86.0%; Pred. No. 2.6e-96;
 Matches 233; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

```

OY 714 CTDNVTDLRIPRGESEGSSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDM 773
Db 1 CTDNVTDLRIPRGESEGSSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDM 60
OY 774 TLDGTSFAPLPHLSLIGIPDVIFEFYRSDNVYQSSSGRSTTIRVRCSPQKTYPGSILLP 833
Db 61 TLDGTSFAPLPHLSLIGIPDVIFEFYRSDNVYQSSSGRSTTIRVRCSPQKTYPGSILLP 120
OY 834 GTCSDGTDCGCPNFMHFMESAAACPLCSVDYHAIVSSCAVGIQKTYVWREPKLSCGIS 893
Db 121 SMCSDGTDCGCPNFMHFMESAAACPLCSVDYHAIVSSCAVGIQKTYVWREPKLSCGIS 180
OY 894 LPEORTYCKTIDFWLKVGISAGTCAILTVLTCYFWMKNQKLEYKSKLVNNAATLKDC 953
Db 181 LPEORTYCKTIDFWLKVGISAGTCAILTVLTCYFWMKNQKLEYKSKLVNNAATLKDC 240
OY 954 DLPADSCALIMEGEVEDDLIFTSKNHSLGR 984

```

DR Pfam; PF00092; vwa; 1.

DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00895; PENTAXIN.
 DR PRINTS; PR00453; VMEFAD0AIN.
 DR PRODOM; PD002153; Pentaxin; 1.
 DR SMART; SM000032; CCP; 34.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00179; EGF_CA; 9.
 DR SMART; SM00001; EGF_1like; 3.
 DR SMART; SM00327; VMA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 5.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 6.
 DR PROSITE; PS50234; VMEF; 1.
 FT SIGNAL.
 KW SIGNAL.
 SQ SEQUENCE 1 17 POTENTIAL.
 3567 AA; 387391 MW; 8FBA8276E12293B5 CRC64;

Query Match 4.2%; Score 230.5; DB 11; Length 3567;
 Best Local Similarity 20.4%; Pred. No. 4.5e-09;
 Matches 174; Conservative 99; Mismatches 300; Indels 279; Gaps 46;

QY 47 HACKSEYHEHYACDSTGSRMAYVPHRPGCLTSLDPVKGTEGSCNAGEFLD--WK 104
 DB 795 HGKSEEMLYKTRCDMDLFFKSAFETTLGNMVP-----SF-CNDADDIDICRL 845
 QY 105 DOSCKPCAGRYSL-----GTGIRFD-EMD-----ELPHGF-----ASLSAN 140
 DB 846 DLTKKCYEYNNYENGFAIGCGMAGNRIDYSDHFDVVOETPTDYGAKRSSRIKRT 905
 QY 141 MELDSDAESTGNCSTSSKVP--RGDYAFNDECTATLMVAVNLKSGTVNFEYYPDS 198
 DB 906 VPLSDPKIOLINIVASVPLPEERNDTLEENQRLIKTLFTITRLKSLNKEPMYS-- 963
 QY 199 STIFFEYVONDOPNADSRMMKTEKGMEHSEVLNNGNVLWRTTASVTKVKKP 258
 DB 964 ---FOLASETVADNSLE-----TEKAF-----LECR-----PGS 991
 QY 259 VLVNRNIAIT--GVAIT--SECFKPGTYADKOGSFCILPANYS--NKGETSC 308
 DB 992 VLRGRMCVCPLGTSYSLEHSTCESCLMGSYQDEBGOLECKLCPRTHTYELHSSVSEC 1051
 QY 309 H-QCDPDKYSEKSSGSSGVNRPACTDKDYFHTNACDANGCFQIMAKKPKIGSEDLGA 367
 DB 1052 KAQCKOGYVSSSGLETSCSPGLTYQPFGRS-SCLLCPETTTTVK-----RGA 1099
 QY 368 VKLPASGVKTHCPNCPNGEFTNNST--CQPCP--YGSYSNGSDCTRCRPAGTEPAGVGEY 423
 DB 1100 VDISAGV---PCPYGEFGRSGLTPCYPCPRDYQOPNAGKSFCLACP----- 1143
 QY 424 KMMWTLPTNMTYVLSGNEFEYKGMTGWEVAGDH--YTAAGASDNDPMI----- 471
 DB 1144 -FYGTTTITGATSTITDCSSFS--STFSAEESIVLVAFGHSQNKYEVSQVHEFCFL 1198
 QY 472 -----LTLVVGFRPPOSVADTENEKEVARITPVEFTLC-----SVN 508
 DB 1199 NPGHNSGTCOOLGRGYVCLCPGV---TGAKCETDIDECSLPLCNGICRDVOYGCTCE 1255
 QY 509 CELFPMGVNSRTNTPVETWKGSGKQSYIIEENTTTSFTWAFORTTTHFASRKTYND 568
 DB 1256 CSL-----GYSGQ-----ICEEN-----INE 1271
 QY 569 VAKIYSINVTNVMGVASCRPCALASDVGSSTGCPNGY---YIDRSGTCHS--CPP 623
 DB 1272 CISPPLCKMTCTDGLASY-----RGT-CVAKYGVHCEETDNECOSSPCLN 1317
 QY 624 NTILKAQPIGYVA-CVPCGPGTKNNK-----THSLCYNDCTSPRTPTRENYNSALA 677
 DB 1318 NAVCK-DQVGFSCPCRPGLGTRCKENDECLSQCPQNGATCKDG-----A 1363
 QY 678 NTYTLGAGPSEFTSKGLKHYHFTLSL--C-GNCGKMASCTDNVTDLRIPESGSGSKI 734

DB 1364 NSERCOPAGFTGT-----HCELININEQSNPCRNQATCVD-----EL 1401
 QY 735 TAYCOAVIIPREVTGYKAGVSSOP--VSLADRLGVTITDMLDGTTSAPAEFLHESIGI 792
 DB 1402 NSYSCKC---QREFSGHRG-ETEDPSGFNLDPREVSGITGYVLLDGV-----LPTLHA 1449
 QY 793 PDVIFPFRSNDV 804
 DB 1450 ITCAFMKSSDV 1461

RESULT 6
 097444 PRELIMINARY; PRT; 709 AA.
 AC 097444;
 DT 01-MAY-1999 (TREMBLrel. 10; Created)
 DT 01-MAY-1999 (TREMBLrel. 10; last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21; last annotation update)
 DE Variant-specific surface protein 417-4 (Variant-specific surface protein type 4 TSP11/TSA417-1like).
 GN VSP417-4.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AD-1;
 RA Ey P.L., Darby J.M., Mayrhofer G.;
 RT "A new locus (VSP417-4/A-I) belonging to a subfamily of tsaa17-like variant-specific surface protein (VSP) genes in Giardia intestinalis."
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 476-705 FROM N.A.
 RC STRAIN=VARIOUS STRAINS;
 RA Ey P.L., Darby J.M.;
 RT "Identification of a subset of tsaa17-like genes within the variant-specific surface protein (VSP) gene family of Giardia intestinalis."
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB065168; AAD0439.1; -
 DR EMBL; U89153; AAD0504.1; -
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR005127; Giardia_VSP.
 DR Pfam; PF03302; VSP; 2.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00001; EGF_1like; 3.
 DR SMART; SM00261; FU; 5.
 SQ SEQUENCE 709 AA; 72664 MW; 6E71F27D2F367F52 CRC64;

Query Match 3.8%; Score 211; DB 5; Length 709;
 Best Local Similarity 20.7%; Pred. No. 1.9e-08;
 Matches 151; Conservative 65; Mismatches 273; Indels 242; Gaps 36;
 QY 4 PGHSHLSARVNGRTERRIPRLRLMLMAGTAFQYTGTPRLACKSEHYEYTAGDS 63
 DB 99 PGHSLTSSDCDGVCTEAPGYFLNPLRANKDSVY-----SCSDT-----TGFTD 144
 QY 64 TGSRMVAVPHRPGCLTSLDPVKGTEGSCNAGEFLDMK--DOSCKPCAGRY----- 116
 DB 145 SKGTYR-----GVQYCERCDGALTDAAGDGAKKRCRCODKYLKNDT 186
 QY 117 -----SLGTGIRPDEMDELPHGFASISANMELDSDAESTGNGTSSKVPVRODYIAFN 169
 DB 187 CVDKAQCDGSGTNKRVAYADDSGNKCYSCSDNLNGVA---NCDTCSYDQSKRT-- 239
 QY 170 TDECT-ATLMVAVNLKSGTVNFEYYPDSIIIEFFVONDQCPN--ADDSRMKATTE 225
 DB 240 --KCTKCTDNNVLTKTSTSGT-----SCVQKQCKDGDGFPKXDS----- 275
 QY 226 KGMFHVSEVLNNGNVLWRTTASVTKVKKPVLYNRILAI-----TGVAYVSECPPC 278
 DB 276 -----SAGKCKLCPNDSTDG-----IANCATCALVSGRSGALVT--CSAC 314

KW Lectin. 1101 AA: 119512 MW: CGB6F5CDBDE656AEC CRC64;
 SQ SEQUENCE

Query Match 3.88; Score 208.5; DB 5; Length 1101;
 Best Local Similarity 18.5%; Pred. No. 5,6e-08;
 Matches 229; Conservative 113; Mismatches 396; Indels 501; Gaps 61;

QY 57 EYACDSTGSRM-RVAVPH---TPGLCTSLP----- 83
 DB 14 DYTDADKLIGKEPREAVPHCAVSANCACISCTGYELTTGANNKTCITLKEDCKTAFSYX 73
 QY 84 DPVAKGT--EGSFSCNAGEF-----LDMKDGSCPKCAGGRSL 118
 DB 74 DKINSTPCKCTCYVNGKEVNTSSHSNDKCYCKNNVNICESLMLKKDKSCGCIIGMSTT 133
 QY 119 GTGIREDEWDELPHGFASLSANMELDDSAESTG---NCTSKWVPKRDYIAFNTDECTA 175
 DB 134 VDG-----SKLCDNATEDNAENKCVGLLASTSSK-----TCDKCFG 170
 QY 176 TLMYAVN---LKSGTVAFERYYPDSSTIEFFVQNDQCPNADDSRWKMTTEKGWEFH 231
 DB 171 --MYSLGSGKCTQKNDKIN-----KCILOVENSQCADGYSLSDKSKAKPE-H 219
 QY 232 SVELKNGNNVLYWRTTAAFSVTKVPEVLYNIAITGVAITSECPCKPGTYADKOGSSF 291
 DB 220 CSKIN-GN-----OCLTMEGYILSKTDSK- 243
 QY 292 CKLCPANSYKNGETSCQCPDKYSEKSSSCNVRACTDKDYFHTACDANGETOQM 351
 DB 244 CTICVDPNPNLSRG--NEC--SIYNAEHCSTCKNR--CTVSGVCYKNNKRCFSPTE-- 295
 QY 352 YKMAKPKICSDLEGAVKLPASGVKTHCPNCPNGEFTKNTSTQCPQYGSY--SNGSDC- 408
 DB 296 --ENKCTKCDNGYFLITTSIGTSPNLYGCFKXANRTECE 331
 QY 409 -----TRCPACTEPVAFEXKWNATLPTNMTETVLGIN----- 442
 DB 332 NGYLENDGKKRCSLCPDTECLTSKTPVPQ-----KLNLRSHLSTGPKCL 382
 QY 443 -----FEYKGTGWEVAGDHIY-----TAGASDNPMLITLVPGRFPQSVW 486
 DB 383 PGLCLSDDDTTCYKCEENGLTNGTCHYFDRKSVLGTSGNHQQCKM--RQYDQYEOLY 440
 QY 487 ADENKEVARITPFEFTLCSVNCCLYFMVGVNSRTNPTVEWKSQKQSYTIIEENTT 546
 DB 441 --NAFKASDNTYTCPLDLXLPYFYSY-----TKG-----TSDNTI 474
 QY 547 TSETMAFOTTEHESARKYTNDAKITYSLNVTNVMNGVASYCRCALASVSGSTSCP 606
 DB 475 TIGCVQGLNVSNDCE--CNDKHIPISIDKASDCVSTITTKLPSCERTAN--GNICTQCP 529
 QY 607 AG-----YIYDSDGTCSCPPNTILKAHOPYGVQ--ACVPC-----GPGTK 646
 DB 530 VQSHGKQKSCGDAHYFDKN--VCKKCPASCSSCSYDSKSKVYVSECEYENIQVYTR 588
 QY 647 NKKHSLCYND-----CTFSRMTPTRTFNYPNS-----A 675
 DB 589 NKENACACINDYKRGPAEDKKKSCAQLNNCKKBEKTEIISGFYVCLDDCDSAYIVGS 648
 QY 676 LANVTYLAGPSF-----TSKGLKYFH-----HFTSLICNGQRKMS 712
 DB 649 QVGAQTQCSPAFENKNNKQOLCSTKQSYGHCACSAFACIJCEDINILITGE--KPCPT 706
 QY 713 VCTD-----NVTDLRIPEGESGFSKSTAYVQAVIIPPEVYIAGVYSSQPSVLAIDL 766
 DB 707 VCKDGFYOIENATD-----GYVCSPC-----PAKCKCTKYNTTSKKVCEV-- 746
 QY 767 IGVTTDNTLGITS---PAELFHLESIGIDPVIFFYRSNDVOTSCS-----809
 DB 747 --ICTEORLDKIRAPACACPTGTQVLE-----NCTQSCSDLSKYPCKCKTD 791
 QY 810 -----SGRS-----TLIRVCSQPKTYVPSGLLPG-----T 835

DB 792 SCNWDSTRTGLIYATECSDGFSGRSPYSNCTT-----CTKSNYTPKEGKNCACODDKCAT 847
 QY 836 CSD-----CTGDCGNFHLW-----ESAACPLCSVADYHA 866
 DB 848 CSDKDTGLTCADPLKVGSKCDGCKTGYMSNECKPCTNHCESSAABCTVCESDYTKV 907
 QY 867 I-----VSSCYAG-----IQKTYVMREP-KLCSG-----GISLEQRYV--TIC 902
 DB 908 ISGNGCNSCVDFYFDEIKGTCTPCTSPCTKCVYKACDEBETCNSEKKTVEECTKG 967
 QY 903 KTIIDFWLKVISAGCTTALLIPLVCYFW-----KKNOLEKYSKLWM-----ATLK 951
 DB 968 STRDHIAEVPVN-GAC-----VCAIYGVGTSTEDNKIECQACKAKVNEFDCSCSK 1018
 QY 952 DCDLPNADSCALWEGEDVEDDLIFTSK--NHSIGRSNHP 989
 DB 1019 DCLRCNAEYLEAKGCEVCVEGYTSSWGSCIPCGRNHP 1057

RESULT 9

Q9XTJ7 PRELIMINARY; PRT: 709 AA.
 AC Q9XTJ7.
 DT 01-NOV-1999 (TREMBLrel, 12, Created)
 DT 01-NOV-1999 (TREMBLrel, 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE Variant-specific surface protein (Fragment).
 GN VSP417-6.
 OS Giardia lamblia (Giardia intestinalis).
 CC Eukaryota; Diplomonadida; Hexamitidae; giardinae; Giardia.
 NX NCBI_Taxid-5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRIS-136;
 RX MEDLINE-99026095; Pubmed-9806870;
 RA Ey P.L., Darby J.M.;
 RT "Giardia intestinalis: conservation of the variant-specific surface protein VSP417-1 (TSA417) and identification of a divergent homologue encoded at a duplicated locus in genetic group II isolates."
 RL Exp. Parasitol. 90:250-261(1998).
 DR EMBL: U08266; AAD03483.1;
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR005127; Giardia_VSP.
 DR InterPro: IPR001239; Kazal_inhib.
 DR Pfam: PF03302; VSP_2.
 DR PRINTS: PR00290; KAZALINHTR.
 DR SMART: SM00261; FU; 3.
 FT NON_TER
 SQ SEQUENCE 709 AA: 71516 MW: 35128B844B38D134 CRC64;

Query Match 3.88; Score 207; DB 5; Length 709;
 Best Local Similarity 20.1%; Pred. No. 4e-08;
 Matches 149; Conservative 70; Mismatches 280; Indels 244; Gaps 35;

QY 60 ACDSTGSRWRAVPHPTDELCTSLDPYKGTCECSFSCA-----GEPLDM----- 103
 DB 18 ACQADGS-----GSAGSCKTCGVAIGQYCECGANAYAVNQCADVNAEGPSKTL 70
 QY 104 --KDQ--SCKPC-----AEGRSIGTGIRFEMDELPHGFASLSANMELDDSAESTG 152
 DB 71 CPARDQCKTCQGGNSMLAGGCYSSGEG-----LGHSLICLSSG-----GDG 113
 QY 153 NCTSSKWPVPGDYIAFNTDECTATLMYAVNLKQSGTVNFEYRY-----POSIIFEEFYOND 209
 DB 114 VCTEAPAGYFLNPLRANTKDSVYSCSDTAGFTDSKTYGQYTCERCBAVLTDAAGDA 173
 QY 210 QCPNADDSRWKMTT-----EKGEHFSVELNNGNNVLYWRTTAAFSVMTKVPVLYR 262
 DB 174 KC--TRGEMKTYLATITGCEGCTPTFEESKEDSDNG-----KRCFAC 214
 QY 263 NAITGVAITSECFCKPCTYADKOGSSFCRLCPANSY---SNGGTSCHO---CDPDKY 316
 DB 215 GDVTTGAVASCERCTPPSP-----DQAKPACIKCGGNNTLKTAAADGTYTCAEQSACSPDSF 269


```

Oy 317 SEKSSSSCN-----VRPACTDKDY-----FYHTAC-----342
Db 270 EVENSQSNRCVLCGDAANGVDKCACTPADGRAAPVTCCTACTDYGKPSADKTCEA 329
Oy 343 -----DANGER-----OLMTKAKPKICSDLEGAUKL---370
Db 330 VSSCTPCCKAKCSNEKENEVCYCTDGDSTYLPFTSOCIOSICASAVNGYALGSAKLCKE 389
Oy 371 -PASVKT-----HCPPNPFEEFTNNSTOCPCPYG-----SYNGSDCTRCAPTEPAVG 420
Db 390 CTAANCKTCDDGRCQCTSDGFYK--NGDACSPECHESKCTKSAGTADCTCECPTG-----442
Oy 421 FEYKMMNTLPMTETTVLSGINFEXKGTGWEVAGDHITAAAGASDNDEPILITLVYPRGR 480
Db 443 -----KALRYGDDGFKGTGAG-----CATOGSGACKTCGLTIDGAS 480
Oy 481 -----PPQSVADTEBK--EVARITFEETLCSVNCLEYEW--GVNSRNTNTPY 525
Db 481 YSECATATEYEPONGVCAKASRATPTCNDSPIONGVCGICANVFEMNGCYETVYKDG 540
Oy 526 ETW-----KSGKQSYTYIIEENTTSFTWAFQRTTEHASKRYNDVAKIYINVTN 579
Db 541 KTYCISAPNGCTQAKADYKIDSGTL-----TVCSBGCEKTS-----STDCIT 585
Oy 580 VMNGV--ASYCRPCALE--ASDVSSCTSCPAGY-IDRDSCTCHSCPPNTILKAHP 632
Db 586 CLDGVKASACTKCDPSCETCGAATTCACATGYITAGSEGACISCESD-----SNGV 641
Oy 633 YGVQACVPCGPGTKNKKHSLCY 655
Db 642 TGKGLNCAPPSNTG-SVLCY 663

RESULT 10
O90ZM3 PRELIMINARY; PRT; 1007 AA.
AC 090ZM3.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Laminin gamma 1 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S., Balasubramani M., Bler M.E.;
RT "Aberrant histogenesis after temporary disruption of the retinal basal
RT lamina."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF373841; AAK55397.1; -.
DR InterPro: IPR004089; Chmtaxis_transd.
DR InterPro: IPR000361; EGF-like.
DR InterPro: IPR000344; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001886; LamNT.
DR Pfam: PF00052; Laminin_B.1.
DR Pfam: PF00053; Laminin_EGF.10.
DR Pfam: PF00055; Laminin_Nterm.1.
DR PROSITE: PS00022; EGF_1; UNKNOWN.6.
DR PROSITE: PS01186; EGF_2; UNKNOWN.6.
DR PROSITE: PS01248; LAMININ_TYFE_EGF; UNKNOWN.9.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1007 AA; 110999 MW; CCBFD9659EF93JPC CRC64;

Query Match 3.7%; Score 206; DB 13; Length 1007;
Best Local Similarity 20.6%; Pred. No. 7.8e-08;
Matches 165; Conservative 80; Mismatches 309; Indels 248; Gaps 46;

```

```

Oy 209 DQCPNADDSRMK--TTEKGEFHSVELNNGNNVLYWRTTAFSVTKRPLYVRIATIT 267
Db 72 EKLPFFNDPRKRAFTASANECLPCDNGRSDECTV-----DPELYRS---T 116
Oy 268 GVAYSSECPCK---PGTYADKOGSSP-----CKLCRANYSK---GETS--307
Db 117 G-HGCHMGCGNDNDGAHCERCNRDSFRLGSEGCJPCSCNPVGSISTQDSTYGCCSK 174
Oy 308 -----CHQCDP--DKYSEKSSSCNVRPACTDKDYFTTADANGETOLMYKAKRK 358
Db 175 PGVMEKCDRCOPGFHSLSEAGCRPCSCNPA-----GSIGECMMETGRC-----218
Oy 359 ICSEDEGAVKLPASGVKTHCPKCPNDFE--KTNNSTOCPCPYGSTSKNSDCTRCAGT 415
Db 219 TCKDNVEG-----FHCRCKRPGFHLDPSPNRCPTPC--FCGHSSVCTN-----261
Oy 416 EPVAFEYKMMNTLPMTETTVLSGINFEXKGTGWEVA--GDIITYAAGASDNDEPIL 472
Db 262 --AVGYSV-----YSITSFQF--GEDEHABQBDGSGVPLQWSEETQDISVI 305
Oy 473 T-----LVVPGRRPOSVMADTENKEVARITFEET--LCVNCLEYFMVGVNSRTN 522
Db 306 SDSYFPIYVAPKFKLGNVLSYGON-----LTFSPVRDRDTRLASADLVLEGAGLRVS 360
Oy 523 TPVETWKSQKQOS-VYTYIIEENTTSFTWA-----FO-----RTTFHEASR 563
Db 361 VPLAOGNSYSESPSLYTFRLHEADYPPRALSAFDOKLHNITAIKIRGYSEKSA 420
Oy 564 KYTNDVAKIYINVTNMGV--ASYCRPCALEASDVSSCTSCPAGYIDRDSCTCHS 620
Db 421 GHLDDV-----TTSVAPGAGVPAAMVEEC-----SCPAGY--EEOFCER 458
Oy 621 CEPNTILKAHPYGVQACVPCGPGTKNKKHSLCY-----LCVNDP--TFSRNTPTRTF 669
Db 459 CSPG--YRREP-GLGPRYSPCVPTCNG--HSETCPDETVC--DCRDNTAGSHCKECS 511
Oy 670 NTFESALNTV-----TLAGSPFTSKGLYFHHFTLSLCGNQGRKMSVCTDNTDLRI 723
Db 512 GYYGDATAGTALDQPCPCPGGSCAVVPRTKEVVCISCTGTGTRKCELCDD--AYFGD 569
Oy 724 PEGSEGFSSKITAYVCOAVIIPPEVTGYKAGVSSQPSIADRLIGVTTDMTLGI-----778
Db 570 PLGENAVAPRCLOCND-NIDPNAVGNCRQTECLKIYNTAGFYCDKCKRQGFGNPL 628
Oy 779 -TSPALFIHESLIGIPDVFFFRSNDVYOSCSGSHSTIRVCSPOKTVPGSILLPETCS 837
Db 629 ATNPA-----DKCHAC-----HCNPYGTVMQOT--SCN 654
Oy 838 D--GTCDGCFHFLWESAACPL-----GSVADYHAIVSSCAVAGIDKTYVWREP 885
Db 655 QVYQCE-CLSHYTERDCSACERGFNFLOSGRGRCRCHALGST-----NGQCDIWTGQ 708
Oy 886 KLSGGSISPEQRYVTKITDF 907
Db 709 CECOPGVT--GORCDRCENHF 728

RESULT 11
O90U21 PRELIMINARY; PRT; 719 AA.
AC 090U21.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Variant-specific surface protein.
CN VSP417-7.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OC NCBI_TaxID=5741;
RN (1)
RP SEQUENCE FROM N.A.
RA Ey P.L., Darby J.M.;

```

"Comparative analysis of the VSP417 subfamily of variant-specific proteins in Giardia intestinalis.";
Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF189719; AAF04387.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001274; Furin-like.
DR InterPro: IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00261; FU; 4.
SEQUENCE 719 AA; 7388 MW; 83BE706BACE7977 CRC64;

Query Match

Best Local Similarity 3.7%; Score 202.5; DB 5; Length 719;
Matches 133; Conservative 72; Mismatches 242; Indels 229; Gaps 41;

```

QY 39 TGGTGPGLHACKSEHYEYETACDSTGSRMRAVPHITPGLCTSLDPYKGT---ECSEFS 94
DB 128 TAATKPSVISC-----DKTG---LITTAHT-----YKGEFCECSGS 163
QY 95 -----CNAGEFLDMKQSC---KPCAEGRYSLGTGIRFDE---WDELPH 132
DB 164 ELTSGODGAKCTKCGASKYL--KDNACVDNAEAGKGYFGKPPAAGNKCITACTDSG 221
QY 133 GPASISANNELDSSAESTGNCSTSKWYPRGDIY---APNDECTATLMYAVNLKQSGTV 189
DB 222 GTAGACAECHYDSTNTRKATCTCTT---DYLRKKADGTTECVAA---NECDTQ 268
QY 190 NEEYYPDSITIEFFVONDOCPNADSRMMKTEKGMETHSVELNRGNVLYWRTTA 249
DB 269 KCFYVAVNST-----GNKCVSCADGAGLAVGTGDAWK-----GV 303
QY 230 SVMTKVPKPVLRNIALTGAVATSECPCKRG---TYADKQSSP---CRLCPA 297
DB 304 DGCACKIRPADIN-----PTKDCBCKPGEISTDKTKCTSTAPPDPIENCKVC-- 353
QY 298 NSYSKGTSGHOCDDPKYSEKSSSCNVRPA--CTDK-----DYRYTHACANGETOL 350
DB 354 -----SEKRRACEECNSNNY-----LPTRMCTIDCKKIGNIYTT-----SNANML 395
QY 351 MYWNAKPKISEDELEGAVKLPASGVKTHCPKPGPFKTN-----NSTCQPCPYGSYSN 404
DB 396 ICKCAVANKKE-----CENTGI--CKTCDGFGYKSSSECKACDSNCKTCNGT--- 442
QY 405 GSQCTRPACTEPAVGFEYKWMNTLPTNMETVLSGINFYKGMGWGVAADHITTAAGA 464
DB 443 SADCCTCLSG-----AVLKYGNDGTGTGCG---AGCATGTGAGA 478
QY 465 SDNDFMIL-----TLVYVGFRRPQSVNADPENKEVA--RITFEVETLCS--VNCELYF 513
DB 479 CKTCGLIIDGTSCSCCAVETEPPOGVCSTTVRAAATCKAGSAKMCNSCTNGFLRM 538
QY 514 MGVNSRRTVPVETWKGSKGOSYTYIIEENTTSFTWAFORTFHEHSKRYTNDV----- 569
DB 539 NGGCVETTKFP-----GKS-----VCEEASAGDTCQKAPGVH-----LNNDLDTGCS 582
QY 570 --AKIYSIV--TNMNG--VAASYCRPALE--ASDVG--SCYCPAGYIIDRDSGTCH 619
DB 583 PGCKTCTSNVTYACMEGYVKTSDSCACAGACATCTGCTACTCTGTGY--KSGTTCV 640
QY 620 SCPPWTIIKAHQPYGVQACVPCGPGTKNNKTHSLCY 655
DB 641 SCTESNSDKTYT--GVANCAKCAP--PLNNKGSVLYC 673

```

RESULT 12

096J07 PRELIMINARY; PRT; 1316 AA.
AC 096J07;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE CDNA FLJ14964 fis, clone PLACE400581, moderately similar to
DF fibroblastin 1 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RX NCBI_TaxId=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otaki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuno Y., Sasaki N.,
RT "NEDO human cDNA sequencing project."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AF027870; BAB5420.1;
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR003410; Hyalin.
DR InterPro: IPR001759; Pentaxin.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF02494; HXR; 2.
DR Pfam; PF00084; Sush1; 3.
DR ProDom; PD002153; Pentaxin; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_5.
DR PROSITE; PS00022; EGF-1; UNKNOWN_6.
DR PROSITE; PS01186; EGF-2; UNKNOWN_8.
DR PROSITE; PS01187; EGF_CA; UNKNOWN_5.
SEQUENCE 1316 AA; 144524 MW; 80615BBA3A4F00A5 CRC64;

```

Query Match 3.6%; Score 200.5; DB 4; Length 1316;
Best Local Similarity 20.6%; Pred. No. 3.2e-07;
Matches 197; Conservative 98; Mismatches 341; Indels 321; Gaps 52;

```

QY 47 HACKSEHYEYETACDSTGSRMRAVPHITPGLCTSLDPYKGTGSEFSCNAGEFLDMK-- 104
DB 562 HGFSEFMEYTKAARCDTDLMKRSEAFETTLGKVP-----SF-CSDAEDIDWML 612
QY 105 -DOSCKPCABGRYSLGTGIRFDEMDELPHGF--ASLSANNELD--SAESTGNCTS 156
DB 613 ENLTKKCYCLEYNDYENGFAIG-----PGWGAANRLDSTDDLPDLYOETATSIGNKS 667
QY 157 S---KWPGRGY---IAFN-----TDECTATLMYAVNLKQSGTVNFEYYPDSITIE 203
DB 668 SRIKRSAPLSDIKIKLIFNITASVPLPDERNDTL-----E 702
QY 204 FVYONDQCPNADDSKMMKTEKGMETHSVELNRGNVLYWRTTAFSVWTKVPR----- 258
DB 703 WENOQRLLQTLFETINKLKRLINKPWSFOL--ASELIT--ADSNLSGKKASPPCRPG 758
QY 259 -VLVRNIAT--GVAIYSE--CPCKPGTYADKQSSFPCKLPANYS--NKGETS 307
DB 759 SVLRGMCVNCPLGYLYNLEHFTCESCRISGYODEGQLCKICPSGMIFYEHSRIISD 818
QY 308 CH-QCDPPKYSEKSSSCNVRPACDKDYFTHACANSETOLMKWAPKICSDLEB 366
DB 819 CKAQCKQGTGSGSLFECESPLGTYPKFGSRCLSC-----PENTSTVKKRG 866
QY 367 AVKLPAASVYKTHCPKCNB--FEKTNSTCQPCPYGSYNSGD--CTRCP-AGTEPAVAF 421
DB 867 AVNISAGCV-----PCEBGRFSRGLMPCHPCRDYVQPNAGARAFCLACPFGYTPPAG- 920
QY 422 EYKMMNTLPTNMETVLSGINFYKGMGWGVAADHIIYTAAGSNDNFILTLVYVGFRR 481
DB 921 -----SRITTECSSPS-----STFSAAES-----VV-----P 943
QY 482 POSVMADETKKVARITF--VFETLCSVN-----CELYFMVGNVSKRN----- 522
DB 944 PASLGHITKRHEITSQVHFECFENPCNHSNGTQOQLGRGVYCLCPUGY--TGKLCETDIDEC 1002

```


QY 523 TPV-----ETWKGSKGOSYIIIEENTTTSTPTAFOPTTEHSAKRY 565
 Db 1003 SPLCLNNGVCKDLVGEFICECPGYTGOR-----CEEN----- 1036
 QY 566 TNDVAKIYSINTVNMNGVASY-CRPCALASDVGSSTSCPAGYIIDRDSGCHS--CP 622
 Db 1037 INECSSPFLCKKGLICVDGVAGYRC-----TCVAGFVLHCETEVNECOSNPL 1084
 QY 623 PNTLKAHPYGVQACVPCGPGTKNNK-----HSLCYNDCTFSRMTPTFTFNYSALA 677
 Db 1085 NNAVCEDVGGFLCKCPGFLGTGCKVDECLSQPCKNATCDG-----A 1131
 QY 678 NTVTLAGSPFTSKGLKFFHFTLSL-C-GNORRMSVCTDNTVTLRIPEGSEGSFSKI 734
 Db 1132 NSFRLCAAGFTGS-----HCELININCCSNPCRNATCVD-----EL 1169
 QY 735 TAYVQAVIIPPEVTGYAGVSSQPVSLADRLIGVTTMDLTGITSRALFL-HLESIGIP 793
 Db 1170 NSYSCKC---QGFSGKRC-----TOSTGFMDEFEVSGIYVLMGLMLPSLHAL 1218
 QY 794 DVIEFYRSD-----VTOSCGSGRTTIRVCSPOKTVPSLLPCTSGDTCDCGNE 846
 Db 1219 TCTFWKSSDDMNTGTPISYANDNSDMT-----LLT--TDVNG----- 1255
 QY 847 HFLW-----ESAACPLCSVADYH-AIVSSCVAGIOTKTYWREPKLCSGGISL 894
 Db 1256 ---WVLVYNGREKITNCSVNDGRWHLIALTWTSANGIMK---VYIDKLSLDGAGL 1306

RESULT 13

Q8T3A6 PRELIMINARY: PRT: 1045 AA.

ID Q8T3A6
 AC 08T3A6
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Y47H9C.4C protein.
 GN Y47H9C.4C
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_Taxid=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 RP MEDLINE=99069613; PubMed=9851916;
 RX none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RT Science 282:2012-2018(1998).
 DR EMBL: AL032657; CAD27615.1;
 SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

Query Match 3.6%; Score 197.5; DB 5; Length 1045;
 Best Local Similarity 19.9%; Pred. No. 4.1e-07;
 Matches 218; Conservative 83; Mismatches 329; Indels 467; Gaps 65;

QY 39 TGTGPE--LHACKESSEHYEY-----TACDSTGSRMRAVPHTPGLCTSLDPYK 87
 Db 224 SDGNGGFCCLNCKCEGKFGAECKEPCQNGATCDNTNGRC-----ICKS---GYH 271
 QY 88 GTECSFSCNAGEF-----LDMKDSCK-----PCABGRYSLGTG 121
 Db 272 GALECNEGCVGFGSGCTOKDCCLNNOCDSSGECCKICMTGKHCDIGCSRGHF---G 327
 QY 122 IRPEWMDLPHGFASLSANMELDSAAE--STG--NCTSSKWVPRGD-----YIAF 168
 Db 328 LQCKONCPCP-----GLEFSDSNASCAAKTGQCCGSGYKPKCDERKCDAGYXGAD 379
 QY 169 NTDECTA---TLMAVANLKSGTVNFE-YIPDSIIIEFFVYVNDCCQPRADSRMMKT 223

Db 380 CSKTCIVAEHTLMCPN---IGFCCKRGFYGDNC---ELACSDSYGPNCEKO----- 428
 QY 224 TEKGEHSEHVELNRGNVLYWRTAFSVTKYKPKVLYVNLNITAVATYSSECP----- 277
 Db 429 -----AMCDWN-----HASECNPEGTSCV 447
 QY 278 CKRGTYADKQGSFCKLCPANSYNSKNGEFSCHOCDDPKYSEKSSCNVRPACTDKDFY 337
 Db 448 CKRG-----RTGKNCSECPLDFT---GPNCAHOC-----QCNOR----- 479
 QY 338 THPAD-ANGETOPLYMKWPKICSEDELEGAVKLPASGVKTHCPNCPGFEFTNNSTCO- 395
 Db 480 -GVCCGADKCCGCCDGMWCHR-CEH-----HCPA-----DTFGANCKE 516
 QY 396 --PCPYGSYNSGD-----CTRCPAGTEPAVGEYKMMNTLPTNMFTYLSGINFEYKGM 448
 Db 517 RCKCPKGI---GCDPIYECT-CPAG-----LOGANCDIGCP 549
 QY 449 TGMENVGD--HITYAGASNDPMILTLVPGFRPOSVADTENKVARITFEFTLCS 506
 Db 550 EGSYGECKLHCKCVNCKCKETGECT-CQGFPGSDCSTTSKGR-----YGESCE 600
 QY 507 VNCLEYFMVGNRSKNTPEFTWK-----GSKG-----KOSYIIIEENTTTSTPTMAF 553
 Db 601 LSCPC-----SDASCCKOTGKCLCPLTGKGVSCDCKDPNFFGLCOETVPS----- 648
 QY 554 QRTTFHASKRYTNDVAKIYSINTVNMNGVASYCRPCALASDVGSST-SCPAGYIID 612
 Db 649 -----PCASTDPKNGVCISCP-----GSSGICHCHENCAGSYGD 683
 QY 613 RDSGTC-----HSCPNTIILKAHPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRMTPT 667
 Db 684 GCQOVCSGADBGDDPTT-----GECI-CEPG-----YHGKTCSEKCPDG 722
 QY 668 TFNYNFS-----ALANTVTLAGSPFTSKGLKFFHFTLSL-----GNORRMS 712
 Db 723 KYGVCALDCPKKASGSCDHINGLJCIPAGLEG---ALCTRPSAGFWNGCRQYV 776
 QY 713 VCTNNVNDLRIPEGSGFSKSTAYVQAVIIPPEVTGYKRAVSSQPVSLADRLIGVTTD 772
 Db 777 RCTSEYKCKMNOTGECs-----CPA-----GFGQDRCDKCE----- 808
 QY 773 MTLDEITSPAEFLHLESIGIPDYIFETRSNDVYOSCS--SGR-----S 813
 Db 809 ---DGYGPDICIKKCKCG-----TATSSCNVSGACHCHGEGTGEFCHALCPE 854
 QY 814 TTIRVRCSPKRTVPSLLPFCSDGTCDCGCFHFLMESAAACPLCSVADYHAIIVSCVA 873
 Db 855 STEFLGCKSECKPKDG-----CGDG-----YECDAIIGCC-----HVDOMSC-- 890
 QY 874 GIQKTTYWREPKLCSGGISLPEQRTVTKTIDFWLKYGISAGCTAILTLVLCYFWK 933
 Db 891 GKAKQEFELANGASTGLT-----WEPVLLLVALLCGGLIAL--EYRN 933
 QY 934 NOKLEYKSKLYMNMATLTKDCLPADSCAIM--EGEVEDDL-----IFTSKN 979
 Db 934 -----KYOK-----EKDPDPTVYSFHAKPANNDEGRQNPPLYRSQVFPDSQAFSEN 981
 QY 980 HSLGRSNHL--PPRGLL 994
 Db 982 N-----GNHOGGFPNGLL 994

RESULT 14

Q8T3A7 PRELIMINARY: PRT: 1070 AA.

ID Q8T3A7
 AC 08T3A7
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Y47H9C.4b protein.
 GN Y47H9C.4B.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9069613; PubMed=9851916;
 RA none;
 RT "genome sequence of the nematode C. elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; AL032657; CAD27614.1; -;
 SO SEQUENCE 1070 AA; 114180 MW; 75254D0DD5643A5 CRC64;

Query Match 3.6%; Score 197.5; DB 5; Length 1070;
 Best Local Similarity 19.9%; Pred. No. 4.2e-07;
 Matches 218; Conservative 83; Mismatches 329; Indels 467; Gaps 65;

QY 39 TGGTGE--LHAKESRYHYEY-----TACDSTGSRWRAVHTTGLCTSLPDPVK 87
 Db 224 SDGMGGEFCINKKEGKFGAEKFECONCATCDNTNGKC-----ICKS---GYH 271
 QY 88 GTCSFSCNAGEF-----LDMKDSCK-----PCAEGRYSLGTG 121
 Db 272 GALCENECGSGFEGSGCTQKCDCLNNONCDSSGCEKCGMTGKHIDICSGRFR---G 327
 QY 122 IRFDENDELPHGFASLSAMNELLDDSAE---SIG--NCTSSKWVPRGD-----YIAF 168
 Db 328 LCKQKCTCP-----GLFSDSNASCDAKTGQCCEGKGPCKDERKCDADQYGAD 379
 QY 169 NTDECTA---TLMYVNLKSGTVNFE--YYPDSIIIEFVQNDQCPNADDSRWKMT 223
 Db 380 CSKTCVARENLMCAPN---IGFCKCKRGFYGDKC---ELACSDISGPNCEKO----- 428
 QY 224 TEKGMFEHVEILNRGNVNLVMTAFSVTWYKPKVLYNIAITGVAYTSECFP----- 277
 Db 429 -----AMCDWN-----HASECNPFITGSCV 447
 QY 278 CKPGTYADKQSSFCIKCPANSYKKEGHSCHQCPDKYSEKSSGSSCNVRRPACTDKDYFY 337
 Db 448 CKPG---RTGKNCSEPCLDFT---GPNCAHQ-----QCNQR----- 479
 QY 338 THPTACD-ANGETQLMYKNAKPKICSEDEGAVKLPASGVKTHCPPCNPGFXTNNSTCO- 395
 Db 480 -GVGDGADGKQCDRGWTHR-CEH-----HCPA-----DTGANCEK 516
 QY 396 --PCRYGYSYNSD-----CTRCPAGTEPAVGFYKMMNTLPTNMETTVLSGINFEYKGM 448
 Db 517 RCKCKGT--GCDPIGECT-CPAG-----LOGAMCDIGCP 549
 QY 449 TGMWEVAGD--HIYPAAGASDNDFMILTLVPGFRRPQSVADMTENKEVARITTFEFLCS 506
 Db 550 EGSYVPGCKLHKCKVNGKCKDEKTEGECT-CQGFEGSGDCTSCSKGK-----YGESCE 600
 QY 507 VNCLEIFVAVGNSTRNTPVETWK-----GSKG-----KQSYTYIIIEENTTSFTWAF 553
 Db 601 LSCPC-----SDASCCKQTGKCLCPLGTKGVSCDQKCDPNTGFLCQELVIPS----- 648
 QY 554 ORTFHFASRYKTYNDVAKIYISINTNWNNGVASYCRCALASVGSCT-SCPAGYIID 612
 Db 649 -----PCASTIDRPNNGVCLSCP-----GSSGIHCHENCPAGSYGD 683
 QY 613 RDSGTC---HSCPNTILKAHPYVOACVPCGPGTKNNKIKHSLCYNDCTFSRNTPTR 667
 Db 684 GCQOVCSCADHGCDPT-----GECI-CEPG-----YHKTSEKCPDG 722
 QY 668 TFFNNFS-----ALANTVTLAGSPSTSGKLYFHHFTSLC-----GNQGRKMS 712
 Db 723 KYGGCALDCKPCASGSTCHINGICIPAGLEG-----ALCTRPCSAGFMWNGCQVOC 776

QY 713 VCTDNTDLRIPEGSEGSKSTIYAVQAVIIPPEVTGYKAVSOPVSLADRLIGVTTD 772
 Db 777 RCTSEYKQCNAGTGBCS-----CPA-----GGQGRCKDPCE----- 808
 QY 773 MTLDGITSPAELEFHLESLGIPVIFRYRNDVTQSCS---SGR-----S 813
 Db 809 ---DGYGPDCKJCKCKQ-----TATSSCNRYSGACHCHPGTEFCHALCP 854
 QY 814 TTIIRNCSPOKTVPGSLILPGTSGDTCGCFHFLMESAAACPLCSVADYHAIYSQVA 873
 Db 855 STEGLKCSKECPKDG-----CGDG-----YECDAIGCC-----HVDQMSC-- 890
 QY 874 GLOKTTYVREKRLKSGSGLSLEPQRYTICKTIDFWLKVGISAGTCTAILTVLCYFVK 933
 Db 891 GKAKQEFELNAGRSTGLT-----WFFVLLIVALCGGLIAL--FYRN 933
 QY 934 NQLEKYSKILVMNNTLKDCLPADSCAIM---EGEVEDL-----IFTSKN 979
 Db 934 -----KQK-----EKDPDKPTVYFHKARVNDGREGFONPLYSROSVPDSDAFSEN 981
 QY 980 HSLGRSNHL--PPRGL 994
 Db 982 N---GNHGGEPNGL 994

RESULT 15

ID 09XWD6 PRELIMINARY; PRT: 1111 AA.
 AC 09XWD6;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Dublin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Tillery-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163339;
 RA Zhou Z., Hartwig E., Horvitz H.R.;
 RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 Engulfment in C. elegans."
 RL Cell 104:43-56(2001).
 DR EMBL; AL032657; CA21739.1; -;
 DR EMBL; AF332568; AAG6061.1; -;
 DR HSSP; P05106; J1V2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin_EGF.
 Pfam; PF00008; EGF; 9.

